

Article

# Robot Tutoring of Multiplication: Over One-Third Learning Gain for Most, Learning Loss for Some

Johan F. Hoorn <sup>1,\*</sup> , Ivy S. Huang <sup>2</sup>, Elly A. Konijn <sup>3</sup>  and Lars van Buuren <sup>3</sup>

<sup>1</sup> Department of Computing and School of Design, The Hong Kong Polytechnic University, Hong Kong, China

<sup>2</sup> Department of Special Education and Counseling, The Education University of Hong Kong, Hong Kong, China; isihuang@eduhk.hk

<sup>3</sup> Department of Communication Science, Vrije Universiteit, 1081 HV Amsterdam, The Netherlands; e.a.konijn@vu.nl (E.A.K.); l2.van.buuren@student.vu.nl (L.v.B.)

\* Correspondence: johan.f.hoorn@polyu.edu.hk; Tel.: +86-528-2766-4509

**Abstract:** In the design of educational robots, it appears to be undecided as to whether robots should show social behaviors and look human-like or whether such cues are insignificant for learning. We conducted an experiment with different designs of social robots built from the same materials, which is unique in robotics research. The robots rehearsed multiplication tables with primary school children in Hong Kong, which is a user group not easily or often accessed. The results show that affective bonding tendencies may occur but did not significantly contribute to the learning progress of these children, which was perhaps due to the short interaction period. Nonetheless, 5 min of robot tutoring improved their scores by about 30%, while performance dropped only for a few challenged children. We discuss topics, such as teaching language skills, which may be fostered by human likeness in appearance and behaviors; however, for Science, Technology, Engineering, and Mathematics (STEM)-related subjects, the social aspects of robots hardly seem to matter.

**Keywords:** robot tutelage; social robots; multiplication; experience design



**Citation:** Hoorn, J.F.; Huang, I.S.; Konijn, E.A.; van Buuren, L. Robot Tutoring of Multiplication: Over One-Third Learning Gain for Most, Learning Loss for Some. *Robotics* **2021**, *10*, 16. <https://doi.org/10.3390/robotics10010016>

Received: 25 November 2020

Accepted: 11 January 2021

Published: 14 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Due to the current COVID-19 pandemic, learners worldwide have come to rely on online teaching and media applications for their education. Nonetheless, the United Nations fear knowledge deficits, learning losses, and gaps in the learning process as a result of a lack of face-to-face interactions ([1], p. 4, 23). Therefore, the United Nations (UN) have pled for different methods of content delivery, such as hybrid learning that is flexible and quasi-individualized ([1], p. 25): “We should seize the opportunity to find new ways to address the learning crisis and bring about a set of solutions previously considered difficult or impossible to implement” ([1], p. 4). If every child had a robot tutor at home, would this—to some extent—make up for missing out on human interaction?

A few years ago, robot teachers were mere science fiction; however, at present, a number of schools have come to include some form of robot education. This varies from educational programs such as Science, Technology, Engineering, and Mathematics (STEM), in which young children learn to build and program robots (see, e.g., [2,3]), to humanoids that teach children mathematics or language (see, e.g., [4,5]). Multiple studies have shown that robots can be beneficial for learning outcomes. A recent review has pointed out that the appearance, behavior, and different kinds of social roles of the robot may positively (or negatively) affect learning outcomes [6].

It seems that people learn better from instructions forwarded by a social robot than by a tablet with the same programs and voice (e.g., [7]). Pupils apparently learn significantly more from their robotic tutors than from a tablet or no robot at all [8,9].

Common understanding has it that in human–human teaching, warm, social, and personal teachers are more successful in advancing the level of study performance of their pupils (e.g., [10–12]). In human teacher–student relationships, a teacher should not just offer theoretical instructions and correct mistakes but also support students personally while creating a healthy relationship (e.g., [10,13,14]). Hamre and Pianta [15] have emphasized that a positive relationship with a teacher makes a child more willing to take on an academic challenge or work on their social–emotional development.

Many researchers have expected to find that robots that show more personalized, pro-social behaviors also render better learning results (see, e.g., [16–20]). However, robot researchers have attempted various forms of social interaction and communicative behaviors but, as a result, obtained a blend of advantageous and unfavorable effects on learning (see, e.g., [6,21]). It seems that individual differences, such as educational ability levels, are sensitive to the level of a robot’s social behaviors: Certain students seem to flourish with a more neutral approach ([21], p. 6, [22]).

Another aspect affecting the so-far mixed results may be the topic that is taught. Robots (as tutors) are employed more frequently in non-STEM subjects such as language (e.g., [23,24]). In language-related topics, such as vocabulary learning or remembering story lines, social behaviors seem to be more beneficial for learning than neutral styles of teaching. For example, when reading aloud narration from a picture book that features fictional characters, facial expressions were shown to be important in bringing the characters to life, such that the children performed better in terms of story recall and target vocabulary [25]. In teaching vocabulary during a storytelling game, cuddly toy robots that appealed to the child’s oral language skills were more successful than robots that did not [16].

In arithmetic and mathematics teaching, such social aspects may play less of a role (e.g., [22,26]). For STEM-related topics, a robot’s social behaviors, such as greeting, following gaze, motivational feedback, and humanoid appearance, do not seem to matter too much (see, e.g., [12,26]) or may even exert adverse effects (see, e.g., [27]). Moreover, robots appear to be successful at maintenance rehearsal and repeated exercise (see, e.g., [28,29]). In other words, if students are to practice multiplication tables as a kind of remedial teaching, the social behaviors of the robot tutor may be insignificant or even distracting [21,27].

Yet, in the on-screen community of virtual tutors and avatars, researchers have reported positive effects of building rapport while learning STEM. For example, a virtual agent was most successful in supporting STEM learning when it showed rapport behavior [30]. Although learners were not aware of the increased rapport, the agent that showed rapport fostered better performance [30]. Arroyo, Royer, and Park Woolf [31] reported that during basic math operations, their adaptive Wayang Tutoring System—embodied by an affective learning companion—improved the working memory and math fluency (the speed to recover or compute answers) of students.

Considering the theory of affective bonding [32], one would also expect that stronger bonding of the learner with the robot enhances learning performance. The affective bond would be fed by the relevance of the robot to the task (here, learning multiplication) and by the robot’s “affordances” or action possibilities (cf. [33]) to execute that task. On the more affective side, emotional bonding can be nurtured by use of a realistic, human-like embodiment and human-like behaviors (cf. anthropomorphism).

In the design literature, the importance assigned to realistic anthropomorphic design can hardly be overstated (see, e.g., [34,35]). For instance, Moshkina, Trickett, and Trafton [36] reported that more humanlike features in a robot, such as a voice, a face, and gestures, invoked more engagement with its audience. Nonetheless, Li, Rau, and Li [37] suggested that a robot’s appearance may exert different levels of likeability, engagement, trust, and satisfaction, depending on the individual’s cultural background. From their empirical work, Paauwe, Hoorn, Konijn, and Keyson [38] concluded that the perceived realism of a robot’s embodiment played a modest role in intentions to use the robot and feeling engaged with it. In robot design, realism is not always key [39].

As factory machines can hardly be altered and university laboratories lack the funds and equipment to build several versions of one robot by themselves, it is often quite a challenge to compare different hardware designs in robot studies. We solved this issue by using Bioloid robot kits, creating a rather unique ensemble of robots that were comprised of the same materials but were different in design. In this way, we were able to see whether the representational variations of a robot—that is, as an animal, a human being, or “just like a machine”—are conducive to learning arithmetic, avoiding the confounding factor of a different make and style of the apparatus.

Our objective in this paper is to investigate if robots can have beneficial effects on learning arithmetic tasks without worrying too much about social, relational, or anthropomorphic issues, thus facilitating the roll-out of tutoring robots in an inclusive manner and at lower costs. To study the effects of robot tutoring on learning a STEM-related task such as rehearsing multiplication, we varied different forms of human-likeness in the design of the robot (cf. [34]). Our initial hypothesis (H1) was that—similar to most of the research community—we expected positive effects of a more humanlike design on rehearsing multiplication.

As our H2, we presumed that working with a robot tutor would potentially be more beneficial for lower-ability pupils than for advanced students. For below-average students, larger progress may be achieved, whereas the added value may be minimal for the high performers.

From Konijn and Hoorn [32,40], one can infer that robot tutoring improves learning multiplication better when the child emotionally bonds with the robot tutor. Bonding is stimulated when the robot’s design looks and behaves similar to a human and, in the perception of the child, is experienced as high in anthropomorphism, relevance, realism, and affordances. Therefore, H3 supposed that building rapport or establishing an emotional bond with the robot would lead to better task performance, perhaps in a mediating or moderating manner. As a control, we queried the social role that the robot played for these children (cf. [41]) and how appealing (“beautiful”) and new they felt their robot tutors were.

Next, we describe the materials and methods we used, which were followed by statistical analyses of the learning outcomes and experiential factors. We conclude with a discussion of the results and our final conclusions.

## 2. Materials and Methods

### 2.1. Participants and Design

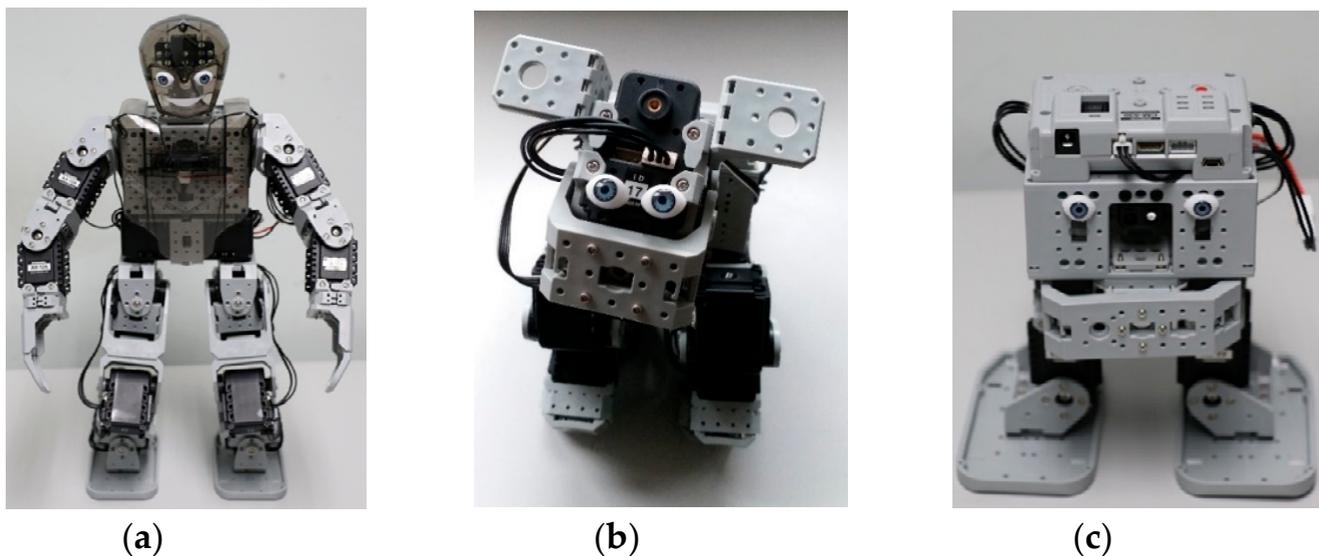
After obtaining approval from the institutional Ethical Review Board, parental consent letters were distributed through two Hong Kong primary schools. Due to strict time planning by the schools and as parents picked up their children early, eventually, 75 students were able to participate in at least one session with a robot tutor and completed the pre- and post-tests ( $N = 75$ ;  $M_{Age} = 8.4$ ,  $SD_{Age} = 0.82$ , range: 7–10, 44% female, Hongkongers). For more details on the study demographics, consult the technical report in Supplementary Materials.

We planned for all pupils to participate in three robot tutoring sessions spread over three weeks (within-subjects). However, due to the tight time schedules of the schools, not every pupil could participate in every session. Children from the S.K.H. Good Shepherd Primary School only participated in one session. This number, plus those from the Free Methodist Bradbury Chun Lei Primary School (who also took one session), resulted in 48 children participating only once. Those who participated twice ( $n = 13$ ), and thrice ( $n = 14$ ) were all from Chun Lei (those participating twice or thrice were different children). For a complete overview of the participatory division, consult the technical report in Supplementary Materials.

To test our hypotheses, we administered an experiment with the between-subjects factors of robot design (3) and advancement level (4), in order to measure their effects on the within-subjects scores, in terms of the multiplication test, before and after robot

tutoring. We also examined the mediating or moderating effects of affective bonding with the robot on learning multiplication. We invited the children to participate in three sessions with the tutoring robot.

The participants ( $N = 75$ ) were randomly distributed over three different robot designs (between-subjects): Humanoid ( $n = 21$ ), Puppy ( $n = 27$ ), and Droid ( $n = 27$ ; see Figure 1). A Chi-square test of independence checked for the distribution of age over robot types, but no significant relationship was found ( $\chi^2_{(6)} = 1.76, p = 0.94$ ).



**Figure 1.** Robotis Bioloids: (<http://www.robotis.us/robotis-premium/>) Humanoid (a), Puppy (b), and Droid (c).

Boys and girls were distributed over the robot design conditions, as follows: Humanoid (15 males, 6 females), Puppy (15 males, 12 females), and Droid (12 males, 15 females). The strict time scheduling of the schools caused an unequal distribution of gender over the three robots; however, this did not result in a significant effect ( $\chi^2_{(2)} = 3.49, p = 0.174$ ).

To determine the advancement level of the pupils, we took the average baseline score ( $N = 75, M = 37.16, SD = 12.88$ ) established in the pre-test and categorized the children into four groups for further exploration. Those who scored lower than one standard deviation below average (baseline  $\leq 22.28$ ) were categorized as “Challenged” students ( $n = 11$ ). Those between one negative standard deviation and the average were categorized as “Below average” ( $22.8 < \text{baseline} \leq 37.16; n = 34$ ). Those between average and one positive standard deviation were categorized as “Above average” ( $37.16 < \text{baseline} \leq 52.04; n = 19$ ), while those beyond one positive standard deviation were categorized as “Advanced” students (baseline  $> 52.04; n = 11$ ). No significant effect of unequal distribution was found between advancement level and robot design ( $\chi^2_{(6)} = 1.73, p = 0.943$ ). For more details, see the technical report in Supplementary Materials.

## 2.2. Procedure

At the Free Methodist Bradbury Chun Lei Primary school, the experiment took place during three weeks on every Tuesday. The S.K.H. Good Shepherd Primary School had time for only one session. In class, the topic and procedure were introduced, and pupils took a 5 min multiplication pre-test consisting of 147 equations (Table 1, Figure 2). One week later, after class, the pupils from Chun Lei were asked to wait in the corridor before entering the experiment classroom (Figure 3).

**Table 1.** Overview of dependent variables.

Variable Name	Description	Number of Items	Abbreviation and Value
<b>Learning variables</b>			
Baseline (from pre-test)	The scores in the pre-test, which established baseline. Pupils multiplied 1- or 2-digit numbers with 2-digit numbers in the range 1–99 (the most difficult equation being $23 \times 67$ ).	147	Base [0, 147]
Final score (from post-test)	Final multiplication score based on multiplying a 1- or 2-digit number with a 2-digit number.	147	FinMSco [0, 147]
Learning gain	Difference between pre-test baseline and post-test final score, which was also calculated as difference-score of FinMSco minus baseline.	147	Fin_min_Base [0, 147]
Gain percentage	The percentage of Fin_min_Base compared with baseline.		Per_Fin_min_Base [min, max]
<b>Experiential variables</b>			
Representation	What does the robot look like to the participant?	3	Human_like = [1, 6] Animal_like = [1, 6] Machine_like = [1, 6]
Social role	What does the robot feel like to the participant?	7	Friend = [1, 6] Classmate = [1, 6] Teacher = [1, 6] Acquaintance = [1, 6] Stranger = [1, 6] Machine = [1, 6] Other = [1, 6]
Bonding	What is the social–affective relationship between the participant and robot tutor? [32,40]	5	Bon_1 ... 5 = [1, 6]
Anthropomorphism	Does the participant attribute human traits or emotions to robot tutor? [32,40]	4	Anth_1 ... 4 = [1, 6]
Perceived realism	Does the robot tutor feel like a real creature or is it a fake? [38,42]	4	Real_1 ... 4 = [1, 6]
Perceived relevance	Is the robot tutor significant for doing the multiplication exercise? [32,40]	4	Rel_1 ... 4 = [1, 6]
Perceived affordances	What can I do with the robot (in view of the multiplication exercise)? [32,40]	4	Aff_1 ... 4 = [1, 6]
Engagement	Level of involvement with the robot	5	Eng_1 ... 5 = [1, 6]
Use intentions	Want to use the robot again?	3	Use_Int_1...3= [1, 6]
<b>Controls</b>			
Novelty	To what extent is the robot tutor new to the participant?	1	Nov_1 = [1, 6]
Aesthetics	To what extent is the robot attractive to the participant in terms of appearance?	1	Aest_1 = [1, 6]
Gender		1	Gender = [Male, Female]
Age		1	Age = [7, 10]

**Figure 2.** Set up for pre- and post-test.



**Figure 3.** Waiting area at Chun Lei.

Those from Good Shepherd were taken out of class one at a time by one of the research assistants and entered the experiment room upon arrival. When one of the pupils of either school entered the room, they were brought by one of the assistants to the table where the robot stood (Figure 4). With the three Bioloid robots available, three children were tutored simultaneously, such that they did not disturb each other.



**Figure 4.** Experimental set-up: Humanoid (left), Puppy (middle), and Droid (right).

The assistant explained that the robot would ask a question and that the pupil could answer through the number pad and pressing Enter (Figure 4). All interactions, tests, and questionnaires were recorded in Cantonese. The robot started the session by asking if the pupil was ready. Upon confirmation, the multiplication program started, automatically drawing 147 equations randomly from various multiplication tables. The equations consisted of one-digit numbers times two-digit numbers (see Table 1). Questioning went on for 5 min, after which the program thanked the child, reported on the number of correct answers, and dismissed the pupil from the session. After one and after two weeks, the same procedure was repeated (at Chun Lei).

The three assistants that operated the robots sat behind a curtain. In this way, the pupil had the illusion that the robot was fully autonomous while, for some functions, someone was pressing buttons on a remote control. The answers that participants typed in on the number pad could be read by the assistant. When the answer was correct, the assistant pressed a button that triggered positive feedback, such as clapping or nodding; when the answer was incorrect, the assistant pressed the button that triggered feedback about the mistake, such as shaking the head or head scratching (對不起。那是不對的。"I am sorry. That is not right").

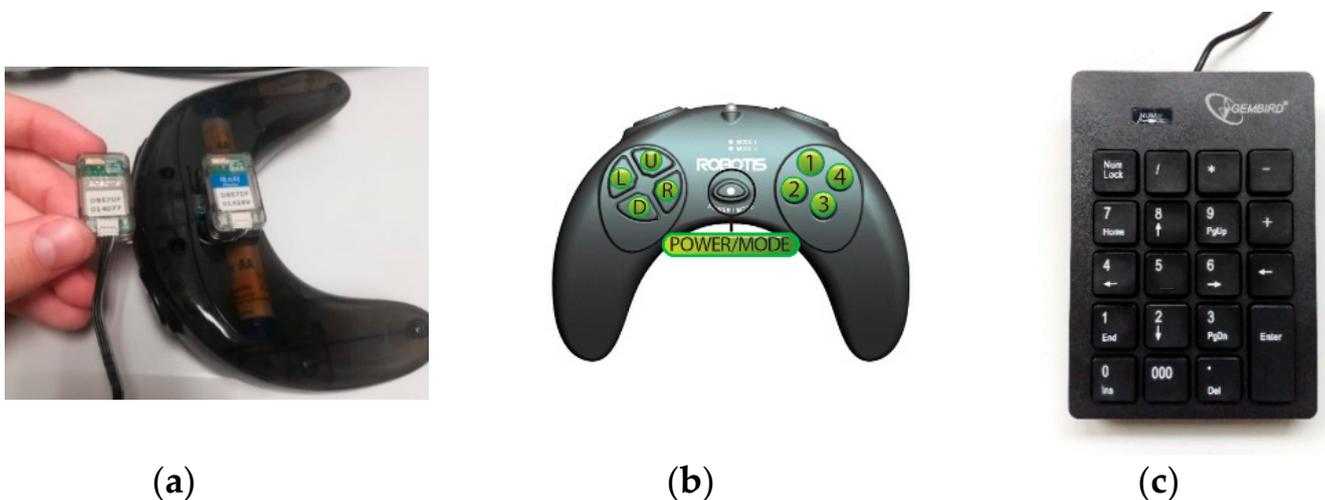
Each time the pupils completed their sessions, they took another multiplication test as a post-test (once, twice, and thrice). The same procedure as in the pre-test was used. After the post-test, the pupils filled out a questionnaire about their experiences with the robot. At Chun Lei, the questionnaire was a homework assignment, while the pupils from the Good Shepherd completed the questionnaire in class.

### 2.3. Apparatus and Materials

The Humanoid, Puppy, and Droid robots (Figure 1) were built from three identical Bioloid Premium DIY kits and programmed on the same CM530 computer (<http://www.robotis.us/robotis-premium/>). To tease out bonding tendencies, we put comparable eyes on the three machines (Figure 1), such that each robot would “look” at the participants. Attached to the Bioloids were Rockbox Cube Fabriq Army (59 × 59 × 59 mm, Bluetooth 4.0, 1 channel mono 3 W) front speakers (Fresh ‘n Rebel, Rotterdam, Netherlands), which were connected to a self-written speech engine in Node.js (a Javascript framework) that ran independently of the robot software.

Trials consisted of pre-recorded Cantonese male speech (23 years of age) of multiplication equations—for instance, “5 times 12?”—and the child’s input was followed by various feedback, such as “I’m sorry, that is incorrect” or “Well done, that’s correct.” Trials were composed from separate audio files of the numbers 1 to 99, of the words “times” and “equals”. Then, the program would randomly select a number audio file, followed by the “times” audio file, followed by another random number audio file, followed by the “equals” audio file.

The speech program kept track of the pupil’s answers, while the motoric functions of the robot were controlled remotely, as the speech program in Node.js was incompatible with the Robotis+ code language of the robot (<https://nodejs.org/en/about/>). Therefore, a wireless Bluetooth receiver was attached to the robot’s computer, which communicated with a wireless controller (Figure 5). The associated code can be found in Supplementary Materials.



**Figure 5.** Wireless Bluetooth receiver (a), wireless controller (b), and number pad (c).

Pupils could input their answers on a numeric keyboard or number pad (OS independent, plug-and-play, 124 × 81 × 21 mm, USB 2.0 powered with type A-plug; see Figure 5) (Gembird, Almere, Netherlands). Apart from audio feedback, a correct answer was rewarded by the Humanoid clapping its hands, the Puppy nodding its head, or the Droid moving up and down. For negative feedback, the Humanoid scratched its head, the Puppy shook its head, and the Droid wiggled from left to right.

The program terminated after 5 min, counted the number of correct answers and, based on the results, played “Well done” or “I’m sorry.” Then, it thanked the child for its participation and asked them to leave the room.

#### 2.4. Measures

Table 1 offers a synopsis of the variables investigated in this study. The full record of variables can be found in Supplementary Materials. Table 1 has two types of dependent measures that are theoretically relevant: learning and experience. Additionally, several control variables are tabulated as well.

**Learning** variables were derived from pre- and post-tests, in which the pupils solved 147 equations drawn from the range [1, 99], with the second number always having two digits (e.g.,  $3 \times 12$  or  $15 \times 31$ ). In the analysis, our main focus is on the Learning gain (the absolute difference between pre- and post-test) and Gain percentage (learning gain relative to a child's baseline knowledge).

We created the measure of Gain percentage because, for example, five more correct answers after robot tutoring may be a relatively big gain for those who performed poorly before but a small gain for those who already performed at a high level (cf. ceiling effect). Then, Percentage\_Fin\_min\_Base was calculated as Fin\_min\_Base divided by the baseline (Table 1).

The **experiential** variables were measured by a 43-item paper-and-pencil structured questionnaire, which was filled out after pupils completed their tutoring session(s) (see Appendix A). Indicative and counter-indicative Likert-type items were scored on a 6-point rating scale (1 = totally disagree, 6 = totally agree). The counter-indicative items on the questionnaire were recoded into new variables, after which we calculated Cronbach's  $\alpha$  for all scales, which was followed by Principal Component Analysis (PCA). From the remaining items, we calculated Cronbach's  $\alpha$  again.

*Representation.* To check the manipulation with the three different robot designs, participants rated to what degree they felt the design of their robot represented a human being, an animal, or a machine. All three dimensions were rated for each robot. In addition, they evaluated the *Social role* of the robot (e.g., a friend or a teacher).

*Bonding* was measured with 5 items (bond, interested, connected, friends, understand). Two examples of indicative items are "I felt a bond with the robot" and "The robot understands me" (Cronbach's  $\alpha = 0.88$ ).

*Anthropomorphism* contained 4 items (machine, human-like voice, human-like reaction, human-like interaction). Two examples are "It felt just like a human was talking to me" and "I reacted to the robot just as I react to a human." Only these two items were left after psychometric analysis by Spearman–Brown correlation ( $r = 0.68, p = 0.000$ ).

*Perceived realism* was based on the studies of [38,42]. This scale had 4 items (real creature, like real, feels fabricated, real conversation), two examples of which are: "The robot resembled a real-life creature" and "It was just like real to me." Psychometric analysis indicated three items for sufficient reliability (Cronbach's  $\alpha = 0.75$ ).

*Perceived relevance* was based on [42] and consisted of four items (important, help, useless, need). Two examples are "The robot was important to do my exercises" and "The robot is what I need to practice the multiplication tables" (with the four items, Cronbach's  $\alpha = 0.73$ ).

*Perceived affordances* was also based on [42] (immediately clear, took a while, puzzled). Two examples are "I understood the task with the robot immediately" and "The robot was clear in its instructions." These two items achieved sufficient reliability ( $r = 0.61, p = 0.000$ ).

*Engagement* was included, in addition to bonding, and was measured based on two scales by [38,42]. Engagement was constructed from 5 items (like, dislike, feeling uncomfortable, fun). Examples are "I like the robot" and "I felt uncomfortable with the robot" (Cronbach's  $\alpha = 0.79$ ).

*Use intentions* were also based on [42]. It consisted of 3 items (use again, another time, help again), an example being "I would use the robot again." These items were deemed only sufficient for group comparisons (Cronbach's  $\alpha = 0.63$ ).

**Control** variables were single items pertaining to novelty ("Have played with robots before"), aesthetics ("The robot looked beautiful"), age, and gender.

### Principal Component Analysis

In the 7- and 5-factor solutions, the divergent validity of the questionnaire items was weak, and the only scale having good measurement quality overall, clearly distinguishable from other components, was bonding (5 items, Cronbach's  $\alpha = 0.88$ ), which was thus the experiential measure used for further analysis. For in-depth PCA analysis, consult Supplementary Materials.

## 3. Results

### 3.1. Preliminary Analyses

Before entering the main analysis, in order to examine our hypotheses, we ran a number of preliminary tests to validate our manipulation and monitor confounding variables, the statistical details of which can be found in the Technical Report of Supplementary Materials. Here, a summary of results will suffice.

We checked the robot design manipulation and found that pupils judged their robots as not significantly different in their machine-likeness; however, the robots were differentiated according to their representation of a human being or an animal. The Humanoid was rated as more human-like and the Puppy was more animal-like; whereas, for the Droid, no significant differences were noted. Thus, all robots were machine-like, with the Droid as the starting point, while the Puppy added an animalistic and the Humanoid a more human-like impression.

We also asked the pupils if they viewed the robot as a classmate, a teacher, a tutor, and other social roles. The different social roles were not significant for human-likeness or animal-likeness, but they *were* significant for machine-likeness ( $F_{(30,246)} = 1.75, p = 0.012$ ), indicating that students perceived a machine-like robot as a machine.

To check for possible confounding effects of non-theoretical variables, we ran several tests of school, gender, and age on performance. Girls carried out more multiplications correctly during the pre-test (but not on the post-test after robot intervention, as we shall see later). The effects of school and gender, while significant on the detailed level (*t*-test), were spurious when more factors were added (*F*-test). Age showed a positive correlation with performance; however, this relation dissolved after robot intervention.

The interaction between advancement level and number of sessions was not significant ( $F = 0.668$ ). More robot-tutoring sessions did not improve learning performance. Notwithstanding that there was not much difference among the groups that took one, two, or three tutorial sessions, we wanted to know how large the learning gain was within each group. We conducted three paired samples *t*-tests of sessions on baseline score versus FinMSco, representing the gain in absolute numbers and in percentages (see Table 2).

**Table 2.** Mean improvement after robot tutoring once ( $N = 75$ ), twice ( $n = 13$ ), or thrice ( $n = 14$ ).

Number of	$M_{Baseline}$	$M_{FinMSco}$	$t$	Sig. (2-tailed)	$M_{Fin\_min\_Base}^a$	$M_{Per\_Fin\_min\_Base}^b$
Sessions = 1	39.71	48.13	$t_{(48)} = -5.66$	0.000	8.42	21.20%
Sessions = 2	35.38	43.06	$t_{(16)} = -3.13$	0.007	7.68	21.70%
Sessions = 3	28.64	39.18	$t_{(11)} = -2.94$	0.015	10.54	36.80%

<sup>a</sup>  $Fin\_min\_Base = FinMSco - Baseline$  <sup>b</sup>  $Per\_Fin\_Min\_Base = Fin\_min\_Base / Baseline$ .

Those who worked once with the robot improved, with 8.42 more answers answered correctly (21.20%). Those who had two sessions had a 7.68 improvement (21.73%) compared to baseline. Those who interacted thrice had a 10.54 improvement (36.83%) compared to baseline. Although the three tutoring sessions seemed to have a better effect, at face value, later in the paper, we see that one-way ANOVA pointed out that the differences among the number of sessions were not statistically significant.

### 3.2. Learning Effects

H1 expected positive effects of robot design on learning, with a significant advantage for Humanoid. H2 assumed differences in learning as a function of advancement level of the students, with the challenged students gaining significantly more from robot tutoring.

To test H1 and H2, we ran a General Linear Model repeated measures of robot design (3)  $\times$  advancement level (4) (between-subjects) on the (within-subjects) number of equations correctly solved before (baseline) and after (final score) robot tutoring ( $N = 75$ ). Note that this was the score in absolute numbers, not the percentage of gain relative to baseline.

Our key finding was a significant and moderately strong main before–after effect on the absolute number of multiplication problems solved correctly ( $V = 0.50$ ,  $F_{(1,63)} = 62.43$ ,  $p = 0.000$ ,  $\eta_p^2 = 0.50$ ). The mean score,  $M_{Final} = 45.73$  ( $SD = 17.40$ ) was significantly larger than  $M_{Baseline} = 37.16$  ( $SD = 14.88$ ) ( $t_{(74)} = 7.19$ ,  $p = 0.000$ ), the mean difference being 8.57 more equations solved correctly after one session of robot tutoring, regardless of robot design or advancement level.

Multivariate tests also showed a significant second-order interaction among robot design, advancement level, and before–after score ( $V = 0.22$ ,  $F_{(6,63)} = 2.99$ ,  $p = 0.012$ ,  $\eta_p^2 = 0.22$ ). Inspection of the mean scores showed that the largest difference was established for Challenged pupils working with the Humanoid ( $M_{Baseline} = 16.33$ ,  $SD = 6.03$ ;  $M_{Final} = 41.67$ ,  $SD = 17.93$ ), while a small reverse effect was found for Advanced pupils working with the Droid ( $M_{Baseline} = 69.33$ ,  $SD = 5.52$ ;  $M_{Final} = 68.00$ ,  $SD = 18.61$ ). However, a paired-samples  $t$ -test showed that the effect for Challenged pupils working with the Humanoid ( $n = 3$ ) was not significant (not even preceding Bonferroni correction;  $t_{(2)} = 3.51$ ,  $p = 0.072$ ), which was probably due to the large  $SD$ s and lack of power. No other main or interaction effects were significant (Supplementary Materials), except for the main effect of advancement level, which was an obviously trivial finding. H1 and H2 were refuted for learning gain in absolute numbers of correctly answered multiplication problems.

#### Learning Gain (Difference Scores)

GLM repeated measures accounts for multiple sources of variance and, therefore, was the strictest test on our hypotheses. To assess if nothing was gained at all from robot design or advancement level, we included fewer sources of variance in our analysis, considering that if lenient tests did not render significant effects either, we could dismiss robot design and advancement level from our theorizing altogether.

Therefore, we calculated the difference score from the final mean score (FinMSco)—baseline score = Final\_minus\_Baseline (Fin\_min\_Base). While 64 pupils gained from robot tutoring, there were 11 (about 15%) who did not perform better, but *worse*, after robot interaction (Fin\_min\_Base =  $-1$  to  $-35$ ). Ten of the worst performers came from the categories Below Average and Challenged, the remaining one coming from the Advanced category. In Figure 6, we show a four-quadrant scatterplot with pre-test baseline as the  $x$ -axis and post-test final score as the  $y$ -axis. The bottom right quadrant contains students who scored high on the pre-test (e.g., 65) but low on the post-test (e.g., 30). The bottom left quadrant has students who did not score too high on either the pre-test (e.g., 10) or the post-test (e.g., 21). These are the students who only learned a little. The top right quadrant contains students who scored high on both the pre-test (e.g., 78) and the post-test (e.g., 79). They too learned a little, but at a higher level. The top left quadrant shows students who scored low on the pre-test (e.g., 17) but high on the post-test (e.g., 51), showing the largest learning gains.

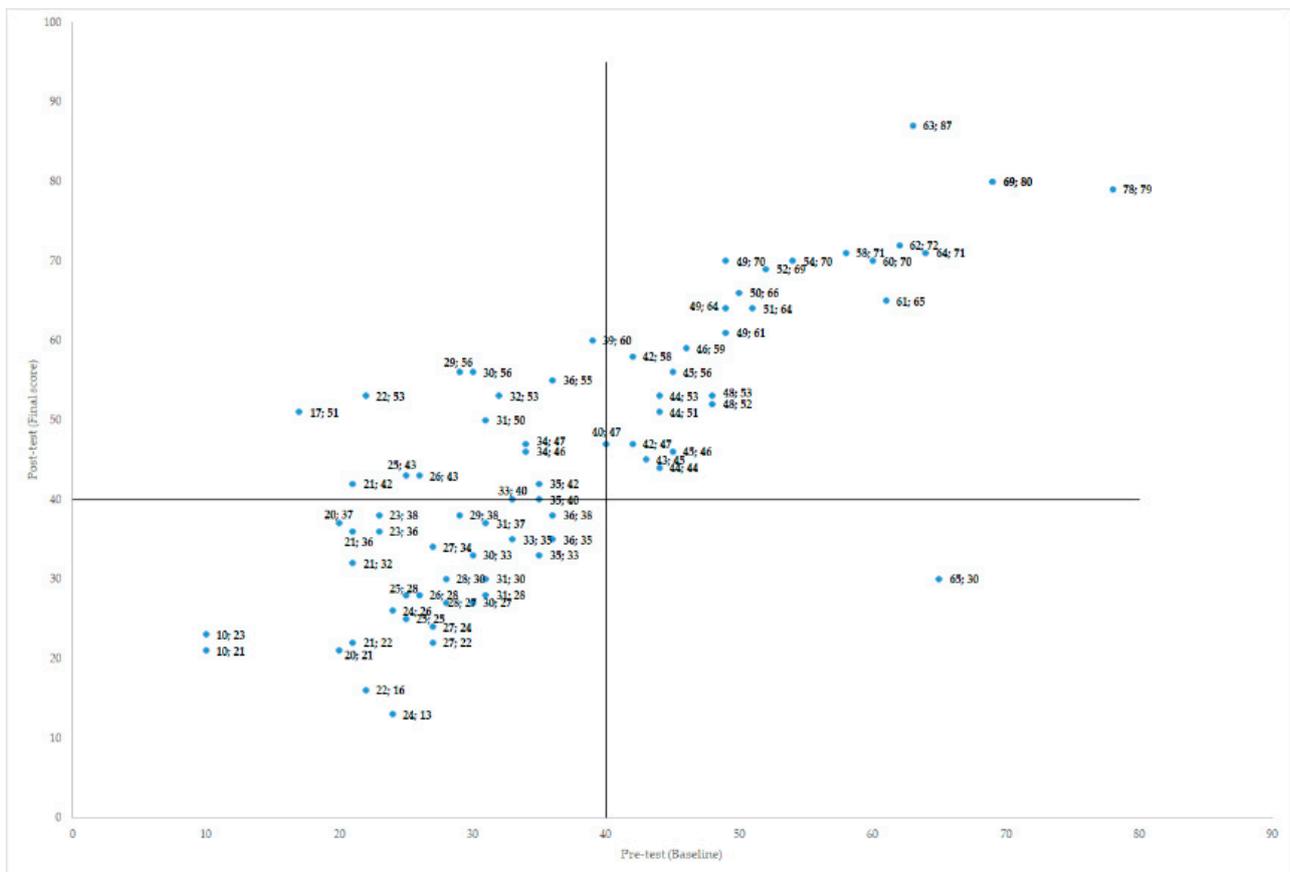


Figure 6. Quadrant analysis for gains and losses (semicolon-separated).

For H1 on Robot Design, we ran a GLM univariate ANOVA of robot design (2) × school (2) × gender (2) on Fin\_min\_Base with age as a covariate (N = 75). The only significant effect was the interaction of robot design × school (2) ( $F_{(2,62)} = 3.33, p = 0.042$ ). Yet, a two-tailed independent samples *t*-test indicated that the main effect of school on Fin\_min\_Base was not significant ( $t_{(73)} = -0.17, p = 0.86$ ). The robot design factor had three levels: Humanoid ( $n = 21, M = 9.47, SD = 1.72$ ), Puppy ( $n = 27, M = 9.50, SD = 1.83$ ), and Droid ( $n = 27, M = 6.81, SD = 1.96$ ). Therefore, we ran three two-tailed independent *t*-tests on Fin\_min\_Base; however, no significant effects were observed (Humanoid–Puppy:  $t_{(46)} = -0.52, p = 0.96$ ; Humanoid–Droid:  $t_{(46)} = 0.84, p = 0.40$ ; Puppy–Droid:  $t_{(52)} = 1.01, p = 0.32$ ). Therefore, neither robot design nor school had a significant effect on learning gains, as measured by Fin\_min\_Base.

We conjectured that, perhaps, certain robot designs exercised negative effects on learning. Therefore, we re-ran the analyses on the group that performed *worse* after robot tutoring. However, robot design and school, again, did not exert significant effects on Fin\_min\_Base. Overall, the effects of schools, gender, and robot designs neither improved nor worsened the children’s learning, as measured through the difference scores.

For the 64 children (about 85%) that did show learning gains after robot intervention, we ran a paired samples *t*-test on baseline versus FinMSco, in order to see *how much* those children gained. The difference between baseline ( $n = 64, M = 37.98, SD = 1.91$ ) and FinMSco ( $n = 64, M = 49.14, SD = 2.05$ ) was highly significant ( $t_{(63)} = -11.20, p = 0.000$ ). On average, those who learned from the robot performed more than one-third better compared to baseline. Although most children learned significantly from robot tutoring, the various robot designs did not significantly differentiate the learning effects, therefore countering H1.

Although robot design did not exact significant effects on learning, perhaps the experience of the design as human-like, animal-like, or machine-like would, allowing yet another chance for H1 to come to expression, albeit in a more perceptual manner. To check the effects of the perceptions of the children with respect to their robot on learning, we carried out regression analyses of human-like, animal-like, and machine-like on *Fin\_min\_Base*. However, no significant relationship was established (human-like:  $t = -0.47$ ,  $p = 0.640$ ; animal-like:  $t = -0.52$ ,  $p = 0.610$ ; machine-like:  $t = -0.50$ ,  $p = 0.620$ ). With gain percentage as dependent (Table 1: *Per\_Fin\_min\_Base*), significant effects remained absent (human-like:  $t = -0.26$ ,  $p = 0.800$ ; animal-like:  $t = -1.16$ ,  $p = 0.250$ ; machine-like:  $t = -0.71$ ,  $p = 0.480$ ).

Combined with the results of the section on learning effects, students perceived the robot as we expected; however, their perception had no effect on learning—not in absolute numbers of correct answers and not as a percentage of improvement from the baseline. Although overall learning gains were achieved, the design of the robot embodiment or what it represented to the children did not matter, thus rejecting H1.

For H2 on advancement level, we ran a one-way ANOVA of advancement level on the difference score *Fin\_min\_Base*, but none of the effects were significant ( $F_{(3,71)} = 1.58$ ,  $p = 0.202$ ). No matter how well or poorly children performed initially, it did not affect their learning gain on average.

As stated under measures, we devised another measure from the notion that children may not have gained differently in absolute numbers, as 8.57 more multiplication problems correct is a relatively stronger gain for a poor performer than for an excellent student. Then, learning gain was calculated using the percentage of gain (*Fin\_min\_Base*) in relation to the baseline ( $\text{Per\_Fin\_min\_Base} = \text{Fin\_min\_Base}/\text{Baseline}$ ). With this measure, we ran a one-way ANOVA of advancement level on *Per\_Fin\_min\_Base* for  $N = 64$ , excluding those with a learning loss. This time, we *did* find significant effects ( $F_{(3,60)} = 12.66$ ,  $p = 0.000$ ) (even with worse performers included, the effect was significant (Supplementary Materials)). On average, the gain percentage (*Per\_Fin\_min\_Base*) increased with the decrease of advancement level ( $r = -0.53$ ,  $p = 0.000$ ) (Advanced:  $n = 10$ ,  $M = 0.17$  (17%),  $SD = 0.11$ ; Above Average:  $n = 19$ ,  $M = 0.22$  (22%),  $SD = 0.14$ ; Below Average:  $n = 25$ ,  $M = 0.35$  (35%),  $SD = 0.28$ ; Challenged:  $n = 10$ ,  $M = 0.90$  (90%),  $SD = 0.61$ ).

To scrutinize the individual contrasts, we carried out six two-tailed independent  $t$ -tests of advancement level with Bonferroni correction (Challenged–Below Average, Challenged–Above Average, Challenged–Advanced, Below Average–Above Average, Below Average–Advanced, Above Average–Advanced) on *Per\_Fin\_min\_Base*. The percentage of learning gain (*Per\_Fin\_min\_Base*) of pupils that were Challenged ( $n = 10$ ,  $M = 0.90$ ,  $SD = 0.61$ ) was significantly higher than those who were Below Average ( $n = 25$ ,  $M = 0.35$ ,  $SD = 0.28$ ), Above Average ( $n = 19$ ,  $M = 0.22$ ,  $SD = 0.14$ ), or Advanced ( $n = 10$ ,  $M = 0.17$ ,  $SD = 0.11$ ) (Challenged–Below Average:  $t_{(33)} = 3.68$ ,  $p = 0.001$ ; Challenged–Above Average:  $t_{(27)} = 4.69$ ,  $p = 0.000$ ; Challenged–Advanced:  $t_{(18)} = 3.73$ ,  $p = 0.002$ ). Yet, the differences among Below Average, Above Average, and Advanced pupils were not significant (see Supplementary Materials). The effects were caused by the Challenged pupils ( $n = 10$ ), indicating that if weak students benefited, they benefited relatively more (90% improvement on baseline) from robot tutoring than others. Calculated as the relative improvement to their individual baselines, H2 could not be rejected for Challenged students, but it could be rejected for the other groups.

### 3.3. Summary of Findings for Learning

1. Prior to robot intervention, pupils performed better with age and girls did better, in terms of baseline performance, than boys. After 5 min of robot interaction, these differences disappeared (main before–after effect on the absolute number of multiplications solved correctly:  $V = 0.50$ ,  $F_{(1,63)} = 62.43$ ,  $p = 0.000$ ,  $\eta_p^2 = 0.50$ ).
2. Most children ( $\approx 85\%$ ) learned from the robot, while a small group ( $\approx 15\%$ ) performed worse (one-way ANOVA of advancement level on percent difference score for  $N = 64$ , excluding pupils with learning loss:  $F_{(3,60)} = 12.66$ ,  $p = 0.000$ ).

3. Those who learned from the robot had an average of more than one-third gain after tutoring (difference between baseline— $M = 37.98$ —and final score— $M = 49.14$ :  $t_{(63)} = -11.20, p = 0.000$ ).
4. The weakest students that gained from robot tutoring did so in percentage of gain (90%), not in absolute numbers, compared to their earlier achievements (significant  $t$ -tests for percent learning gain only with inclusion of Challenged students:  $t_{(33)} = 3.68, p = 0.001$ ;  $t_{(27)} = 4.69, p = 0.000$ ;  $t_{(18)} = 3.73, p = 0.002$ . All other contrasts were not significant)
5. School, gender, design of the robot, the number of times these children were tutored, nor the experience of novelty of the robot were influential for learning through robot tutoring (i.e., none of the control variables had significant effects on learning or they caused trivial findings).

### 3.4. Experience

Although we utilized a range of psychometric scales in our questionnaire to measure different dimensions of affect (i.e., engagement, bonding, anthropomorphism, perceived realism, relevance, perceived affordances, and use intentions), none but bonding achieved convergent *and* divergent measurement reliability (Supplementary Materials). Therefore, we decided to work with the only clear-cut case we had—bonding—and not to make ad hoc decisions.

H3 expected that emotional bonding with the robot would positively affect the learning outcomes in a mediating or moderating way. To examine H3, we once more ran the previous GLM repeated measures of robot design (3)  $\times$  advancement level (4) (between-subjects) on the (within-subjects) number of equations correctly solved before and after robot tutoring, but now with mean bonding as the covariate. However, mean bonding exerted no significant main or interaction effects on the multiplication scores, and the earlier pattern of results was not altered (Supplementary Materials).

To allow the presumed relation between bonding and learning to occur more easily, we ran a two-tailed bivariate correlation analysis between  $M_{Bond}$  and  $Fin\_min\_Base$  ( $r = 0.007, p = 0.951$ ) and between  $M_{Bond}$  and  $Per\_Fin\_min\_Base$  ( $r = -0.076, p = 0.531$ ). Neither were significant.

Therefore, H3 was rejected. Bonding tendencies were independent of the design of the robot or the advancement level of the children. The level of bonding with a robot tutor seemed not to have any substantial correlation with learning, not in absolute numbers nor in relative gain.

To check whether any of the non-theoretical variables affected the level of learning and bonding, we conducted multivariate analysis of robot design, advancement level, school, and gender on  $Fin\_min\_Base$  and  $M_{Bond}$  and on  $Per\_Fin\_min\_Base$  and  $M_{Bond}$  with age, novelty, and aesthetics as covariates. However, the only significant effect that included bonding was that aesthetics covaried with  $M_{Bond}$  ( $F_{(1,71)} = 13.21, p = 0.001$ ); that is, a robot that was experienced as “prettier” raised stronger bonding tendencies. For further statistical details, consult Supplementary Materials.

### Effects on Bonding

We ran a univariate analysis of variance (ANOVA) of robot design and advancement level directly on mean bonding. Not all children who took the multiplication test also filled out the questionnaire; therefore,  $N = 70$ . The intercept was significantly different from zero, such that bonding tendencies did occur ( $F_{(1,58)} = 194.76, p = 0.000, \eta_p^2 = 0.77$ ). However, none of the main effects or interactions were significant ( $F < 1$ ; see Supplementary Materials). Neither robot design nor advancement level exerted significant effects on bonding.

As an extra exploration, we conducted an ANOVA of robot design (3)  $\times$  advancement level (4)  $\times$  school (2)  $\times$  gender (2) on the grand averages of  $M_{Bond}$ , showing that only the difference in school was significant ( $F_{(1,34)} = 4.57, p = 0.04$ ). We ran an independent samples

$t$ -test of school on  $M_{Bond}$ , showing that bonding at good shepherd was significantly higher than at Chun Lei ( $t_{(68)} = 2.99, p = 0.004$ ). Theoretically, this is an irrelevant finding.

We then ran three  $t$ -tests with sessions as the grouping variable (once–twice, once–thrice, and twice–thrice). The effects on  $M_{Bond}$  of once and thrice and that of twice and thrice were not significant (once–thrice:  $t_{(54)} = 1.31, p = 0.20$ ; twice–thrice:  $t_{(20)} = 0.97, p = 0.34$ ). However, the difference between once and twice was significant for  $M_{Bond}$  (once–twice:  $t_{(60)} = 3.01, p = 0.004$ ), even if  $\alpha$  was corrected to 0.017 (with respect to Bonferroni). Apparently, mean bonding was lesser upon second encounter ( $M_{Bond1} = 3.60, SD = 1.64$ ;  $M_{Bond2} = 2.19; SD = 1.70$ ), which was due to Chun Lei pupils alone. The insignificant difference with those encountering the robot thrice might indicate a ceiling effect.

We wondered if the high bonding upon first encounter was due to a novelty effect, wearing off after multiple encounters. Therefore, we correlated  $M_{Bond}$  with Novelty and found that the correlation was significant but not very strong ( $r = 0.31, p = 0.01$ ). Children from Chun Lei saw the robot more often, such that the lesser novelty may have led to lower rates of bonding.  $M_{Bond}$  also correlated with aesthetics ( $r = 0.56, p = 0.000$ ), indicating that the experience of “prettier” led to stronger bonding tendencies, as supported by the covariance analysis above.

### 3.5. Summary of Findings for Experience

With respect to the experience of the robot tutor as a social entity, we found the following:

1. The pupils perceived the robot as intended (manipulation successful; significant  $t$ -tests for ratings on human-likeness and animal-likeness, not on machine-likeness).
2. The social role they attributed to the robots had no significant effect on their perceptions of human-, animal-, or machine-likeness, except that the role of “machine” indeed raised significant machine-likeness, which was a trivial finding (different social roles not significant for human-likeness and animal-likeness, solely for machine-likeness:  $F_{(30,246)} = 1.75, p = 0.012$ ).
3. From a design perspective, the Bioloids, to these children, were basically all machines similar to the Droid, while the Puppy added animal-like features to that basic frame, and the Humanoid added human-like features to it. However, the type of robot (humanoid, animal, or machine) did not affect the bonding tendencies (mean bonding as a covariate did not evoke significant main or interaction effects on the multiplication scores in GLM repeated measures of robot design and advancement level).
4. Only the bonding scale was psychometrically reliable; all other measures for these children seemed to be related to that experience or were confusing (cf. Cronbach’s  $\alpha$  in combination with Principal Component Analysis).
5. Bonding had no significant relation with learning gains. After 5 min of robot training, the children improved their skills regardless of the quality of the established relationship: The bonding intercept was significant ( $F_{(1,58)} = 194.76, p = 0.000, \eta_p^2 = 0.77$ ), but there were no significant effects on learning, see bullet 3.
6. The Good Shepherd children experienced more bonding with their robot tutor than Chun Lei pupils, maybe owing to a novelty effect (trivial finding:  $t_{(68)} = 2.99, p = 0.004$ ).
7. Stronger perceptions of the robot’s attractiveness (“beautiful”) were associated with stronger bonding tendencies (mean bonding correlated significantly with aesthetics:  $r = 0.56, p = 0.000$ )

## 4. Discussion and Conclusions

We found that 5 min of robot tutoring improved the learning of multiplication regardless of the design of the robot or the advancement level of the pupils. This result countered our hypothesis H1: that a more anthropomorphic design would enhance performance. It also countered H2, regarding different effects for advancement level when dealt with as the absolute number of equations solved correctly. However, H2 was not refuted when seen as the relative gain pupils obtained from robot tutoring, as compared to their earlier achievements; then, the more challenged children ( $n = 10$ ) gained relatively more than the others.

H3, which considered that a child learns more while developing a stronger emotional bond with the robot tutor, was also disconfirmed. While rehearsing multiplication equations in this study, learning and bonding seemed to be two different strands of processing, both happening but not affecting each other significantly.

Thus, our conclusion is very straightforward: Apparently, children improved, in terms of their multiplication table performance, after 5 min of exercise with a robot. More sessions were unnecessary. Initial differences between gender, age, or school disadvantages were compensated for, and the novelty of the method had no significant effect on learning. The type of robot or its social role (teacher, peer, friend) did not matter, either (cf. [43]): A more human-like machine did not improve performance, a teacher role was no better than a peer, and the level of emotional bonding of the child with the tutoring machine (e.g., as it is new and beautiful) had no significant effect on their learning outcomes.

This is good news for teaching practice (cf. [1]), as cheap and simple robots of whatever kind may help the larger part of pupils gain more than 33% better scores with little time and financial investment. The weakest pupils should be treated with caution: Many may have a 90% progress, but some challenged and under-average children may be set back by robot tutoring. For different reasons, challenged as well as certain advanced students can be easily distracted and may experience learning difficulties (see, e.g., [44]).

The theory of affective bonding [32,40] was not supported. For the children of the study, the different conceptualizations of affordances, relevance, realism, and anthropomorphism seemed to be diffuse, except for the notion of bonding (“I felt connected to the robot”); such bonding may be present but was not influential for rational performance.

Robots are not human beings (cf. [43]). It may be that a warm relationship with a human teacher makes a child want to work harder and may improve their social–emotional development (e.g., [10,13–15]). In project-based learning, social interaction is important, as it is classroom-oriented and requires the student to actively explore real-world challenges and cases, providing multiple perspectives. Our robot merely helped, one-to-one, with the maintenance rehearsal of arithmetic equations that have one specific answer. For a simple drill such as quickly practicing multiplication with a little robot, warm relationships did not seem to be necessary, perhaps because the interaction was so short. According to Serholt and Barendregt [45], it may be that children do not develop bonds with robots in the human sense but engage in a different sort of relationship; what this relationship is needs further study.

Our work coincides with the results of Hindriks and Liebens [26]: that social behavior during a maths task is not conducive to learning. Moreover, for certain challenged pupils, the effects we found were even counterproductive. It seems that matching the robot’s appearance with its task is insignificant, despite some individual preferences for specific robot appearances in some tasks [21,37,46,47]. Our robots were successful at maintenance rehearsal and repeated exercise (e.g., [28,29]); during the remedial teaching of a strongly rational task, the bonding aspects of the robot appeared to be unimportant.

A strong point of our study was the comparability of the three robot designs. It is quite hard to compare existing factory robots of different makes, telling which design elements are responsible for the differences in user responses. Our basic design, materials, and general appearance of the robots was similar but differentiated in representation: It is a rather unique finding that the children recognized the basic design of all three robots as a machine with human features added for the humanoid and animal characteristics for the puppy. Unexpectedly, these representational variations were not conducive to learning, which brings us to the limitations of this study.

Field studies add to ecological validity and plausibility, yet at the cost of methodological soundness. The time schedules of schools and parents left us with 75 children that could participate in but one session; therefore, the insignificant progress after the second and third session may have been due to a lack of power. Effects of the advancement level (i.e., weaker or stronger pupils) may have also been disturbed by the small numbers in the study. Working with children, in itself, already yields noisier data than with adults, which

may have drowned some effects of taking multiple sessions, the mix-up of psychometric constructs (e.g., anthropomorphism, realism), or the effects on bonding. It may be argued that 5 min of interaction is too short to become attached to a machine. Additionally, our robots were not actually “teaching” but, rather, rehearsing content materials or taking tests. The robot simply gave feedback (correct/incorrect) to the child, and the only social behavior exhibited was the gesture performed.

#### *Future Outlook and Research Directions*

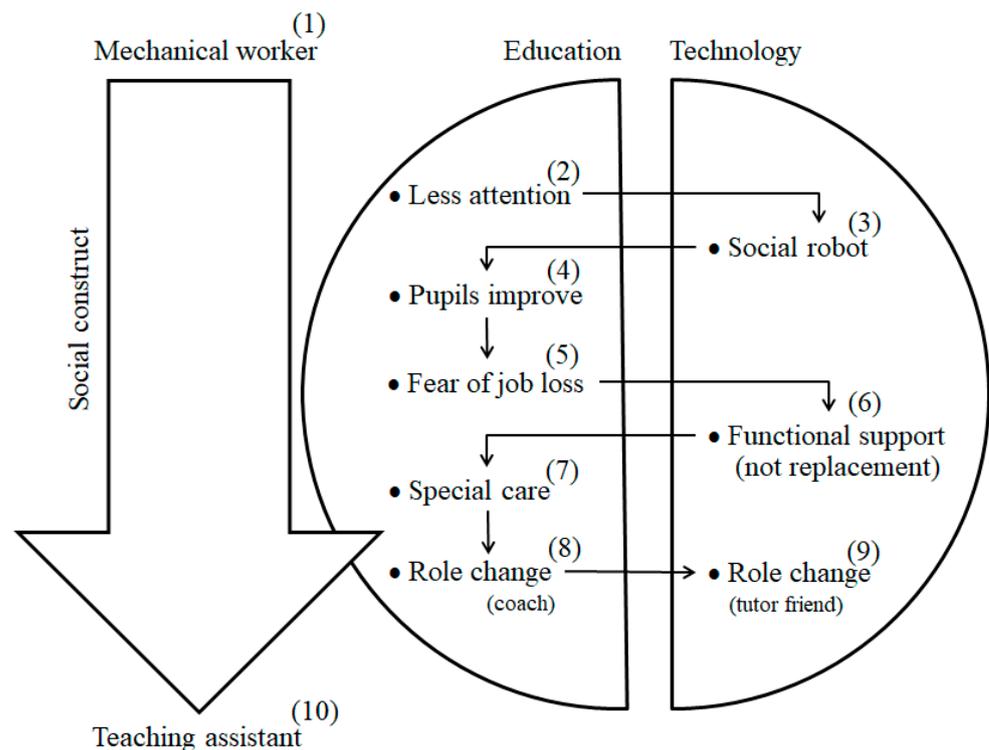
Due to severe budget cuts and fewer teachers, education faces a lack of human resources to serve an increasingly larger number of pupils with a wider variety of individual needs. Owing to changes in care systems (i.e., in Europe), children with special needs are often integrated into regular—rather than special—schools (see, e.g., [48]; for the situation in Hong Kong, see [49]). Migration causes new mixes of children from diverse backgrounds, with cultural and educational differences. The current pandemic has led to a demand for novel teaching solutions in order to make up for learning loss [1]. These transitions demand ways of teaching that differ from class-wise instructions [1]. As is, the teaching level converges to the middle, whereas children learn most if the instructions match their level of proficiency [50].

Social robots may provide support, which probably has far-reaching implications for classroom instruction and organization. For example, repetitive tasks may be performed by the robot, while the teacher focuses on special cases or develops and teaches advanced topics. This actually asks the teachers to recalibrate their profession. In the near future, teachers may have to consider working in teams that consist of synthetic colleagues. However, before the role of these new robot colleagues can be outlined, we must understand how a robot’s (limited) capabilities can match not only the teaching needs of pupils but also those of teachers. In this respect, moral deliberations on robots in education should be proliferated (e.g., [51]).

Our results suggest that a robot does not have to be fancy, in terms of looks or behavior, to help children to increase their performance quickly in arithmetic rehearsal tasks. In this study, weak pupils benefited strongly from robot instruction, with the exception of a few challenged children. Robot teachers in motion pictures and comic books do not have to remain mere science fiction. Educators and parents may apply a simple and cheap machine equipped with the proper software in order to make up for knowledge deficits and gaps in the learning process without having to fear the lack of face-to-face interaction. This makes robot tutoring even more feasible in the context of the COVID-19 pandemic.

Hence, we may consider to scale-up or sustain a STEM education program based on robotics, as children might not be able to attend lessons in classrooms and need to learn from home, through online lessons, during lockdowns. Social robots may be one way to influence and change education, as asked for by the UN [1], beyond its intended use purpose of “robot tutoring” alone, in order to allow for safe learning from home. However, the social construct of a robot still is that of a mechanical worker fit for low-quality jobs (Figure 7, (1)). Then, as a future research direction, we may investigate how the mutual shaping of education and technology will transform the way we teach. Next follows a number of hypotheses to pursue.

The current social norm is still that affective tasks, such as nursing and teaching, should not be left to machinery. However, as increasingly fewer people are available in education and children must stay at home, the real teacher has less and less time, particularly for pupils who need special attention (2). Technologists have offered solutions by developing social robots that can take over (at least, the simple) school lessons, as we saw in our study (3). Indeed, these pupils are moving forward and, however undifferentiated, if they regard the robot as sociable and nice, they develop a positive attitude (4). Teachers observe this and may worry that their jobs are being taken from them (we have heard such stories), which is an initially negative attitude (5).



**Figure 7.** Mutual shaping of robot technology, teachers, pupils, and parents.

In turn, the technologists may now adapt the functionality in such a way that the robot performs supporting tasks and does not replace the teacher (6). Now, the teacher is satisfied that they can pay attention to special cases (7), while the robot carries out the more tedious maintenance rehearsal. Responsibilities become differently distributed. The role of the teacher becomes more focused on individual coaching and less on “mass education” (8). Parents see that students move forward and that their children are happy with their robot (9). Therefore, in society, the social construct of a robot is expected to change from a low-skilled mechanical worker to a kind assistant that can “teach” (10). Moreover, the children that were taught by robots enter society with yet another preconception of robots: as a teacher and as a personal friend but without the moral pitfalls of teacher–pupil friendships. Moreover, these children know from their own experience the “dos and don’ts” of robot tutoring; therefore, some of them may become more sophisticated robot researchers and designers than we are today.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/2218-6581/10/1/16/s1>, Technical Report: Bioloids, multiplication TechRep, Software: Bioloids, code, Audio: Bioloids, audio files.

**Author Contributions:** Conceptualization, E.A.K. and L.v.B.; methodology, E.A.K., L.v.B. and J.F.H.; software, L.v.B.; validation, L.v.B., J.F.H. and I.S.H.; formal analysis, I.S.H. and J.F.H.; investigation, L.v.B. and I.S.H.; resources, J.F.H. and I.S.H.; data curation, J.F.H.; writing—original draft preparation, J.F.H., I.S.H. and L.v.B.; writing—review and editing, J.F.H. and E.A.K.; visualization, J.F.H. and L.v.B.; supervision, J.F.H. and E.A.K.; project administration, J.F.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the FSW Research Ethics Review Committee (RERC) (protocol code lbn344, date of approval: 20 April 2019).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data can be made available by the corresponding author.

**Acknowledgments:** We would like to thank the pupils, parents, and staff of the Free Methodist Bradbury Chun Lei Primary school (especially principal Liu) and of the S.K.H. Good Shepherd Primary school (especially principal Chan) in the Hong Kong SAR for their kind co-operation in the experiments. Tom Ootes and Prashant Sridhar are kindly acknowledged for their help in programming the software. Ruowei Li (Tristan) is kindly acknowledged for assembling two of the Bioloid robots. Without the hands-on support of Stefanie de Boer, Marijn Hagenaar, Eddie Hong, Kelvin Lai, Gigi Lee, Edison Liang, Dominique Roos, and Dickson Soon, we would not have been able to conduct the experiments. The anonymous reviewers are greatly thanked for their helpful comments.

**Conflicts of Interest:** The authors declare that they have no conflict of interest.

## Appendix A

Structured questionnaire on the experience of a tutoring robot (English translated from Cantonese). Variable names (between brackets) were left out from the original questionnaire.

What did the robot look like to you? The more circles you fill in, the more you agree with the statement. Only one circle filled in means you don't agree at all, all circles filled in means you totally agree.

機器人對你來說像什麼呢？你填滿越多的圈圈代表你越認同對應的陳述。只填滿一個圈圈代表你完全不同意，如果所有圓圈都被你填滿了，代表你十分認同這個陳述。

### [Representation]

The robot looked like a ...

機器人看起來像 ...

1. Machine  
機器
2. Human  
人類
3. Animal  
動物

### [Social Role]

What did the robot feel like to you? To me the robot felt like a ...

(choose one answer that suits you best)

你怎麼看待機器人呢？對我來說，機器人像一個 ...

(選擇一個最接近你想法的)

4. Friend  
朋友
5. Classmate  
同學
6. Teacher  
老師
7. Acquaintance  
熟人

8. Stranger  
陌生人
9. Machine  
機器
10. Other ...  
其它 ...

How did you feel about your connection with the robot? The more circles you fill in the more you agree with the statement.

你覺得你跟機器人的關係怎麼樣呢？越多的圈圈代表你越認同對應的陳述。

### [Engagement]

The robot ...

這個機器人 ...

11. I like the robot  
我喜欢機器人
12. The robot gave me a good feeling  
它讓我感覺很好
13. I felt uncomfortable with the robot  
機器人令我感覺不舒服
14. It was fun with the robot  
機器人好好玩
15. I dislike the robot  
我不喜欢機器人

### [Bonding]

16. I felt a bond with the robot  
我覺得和機器人有連結
17. I felt like the robot was interested in me  
我覺得機器人對我有興趣
18. I felt connected to the robot  
我對機器人有連結的感覺
19. I want to be friends with the robot  
我想和機器人做朋友

20. The robot understands me

機器人明白我

What did you think about your interaction with the robot? The more circles you fill in the more you agree with the statement.

你覺得你跟機器人的互動怎麼樣？越多的圓圈代表你越同意。

#### [Anthropomorphism]

21. To me the robot was a machine

我覺得機器人只是一個物件

22. It felt just like a human was talking to me

我覺得好像一個人和我說話

23. I reacted to the robot just as I react to a human

我跟機器人對話猶如和人類對話一樣

24. It differed from a human-like interaction

和機器人交流和人類不一樣

#### [Perceived Realism]

25. The robot resembled a real-life creature

機器人猶如活物一樣

26. It was just like real to me

機器人好真實

27. The robot was fabricated

機器人好做作

28. It felt just like a real conversation

和機器人對話好真實

#### [Relevance]

29. The robot was important to do my exercises

機器人對我學習很重要

30. The robot helped me to practice the multiplication tables

機器人幫到我練習乘法表

31. The robot was useless for rehearsing the multiplication tables

機器人幫不到我練習乘法表

32. The robot is what I need to practice the multiplication tables  
 我需要機器人才會練習乘法表

#### [Perceived Affordances]

33. I understood the task with the robot immediately  
 我明白機器人的指示
34. The robot was clear in its instructions  
 機器人的指示好清晰
35. It took me a while before I understood what to do with the robot  
 我需要一點時間明白機器人的操作
36. I puzzled to understand how to work with the robot  
 我對於機器人的用法有點疑問

#### [Use Intentions]

For the next time practicing multiplications, I would ...  
 下次練習乘法表的時候，我會 ...

37. use the robot again  
 再次用機器人
38. use another tool, like a tablet  
 使用其他學習工具
39. want this robot to help me again  
 想要機器人再次幫我

Then, some final questions

*The more circles you fill in the more you agree with the statement.*  
 最後幾個問題，圈得越多代表你越同意。

#### [Novelty]

40. I played with robots before  
 我有玩過機器人

#### [Aesthetics]

The robot looked ...  
 機器人的外表 ...

41. Beautiful  
 很漂亮

#### [Demographics]

42. I am a ...  
 我是一個 ...

- Boy 男孩
  - Girl 女孩
43. How old are you 請問你幾歲？
- 

**Thank you for all the help. See you next time!**  
 謝謝你的幫助。期待我們下次再見。

## References

- UN. Policy Brief: Education during COVID-19 and Beyond. United Nations. Available online: [https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/08/sg\\_policy\\_brief\\_covid-19\\_and\\_education\\_august\\_2020.pdf](https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/08/sg_policy_brief_covid-19_and_education_august_2020.pdf) (accessed on 30 August 2020).
- Gomoll, A.; Šabanović, S.; Tolar, E.; Hmelo-Silver, C.E.; Francisco, M.; Lawlor, O. Between the social and the technical: Negotiation of human-centered robotics design in a middle school classroom. *Int. J. Soc. Robot.* **2017**, *10*, 309–324. [[CrossRef](#)]
- STEMex. STEMex Learning Centre. 2019. Available online: <https://stemex.org/about/> (accessed on 9 December 2019).
- Chang, C.W.; Lee, J.H.; Chao, P.Y.; Wang, C.Y.; Chen, G.D. Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *J. Educ. Technol. Soc.* **2010**, *13*, 13–24.
- Nuse, I.P. Humanoid Robot Takes Over as Teacher. Available online: <http://sciencenordic.com/humanoid-robot-takes-over-teacher> (accessed on 30 September 2017).
- Belpaeme, T.; Kennedy, J.; Ramachandran, A.; Scassellati, B.; Tanaka, F. Social robots for education: A review. *Sci. Robot.* **2018**, *3*, eaat5954. [[CrossRef](#)] [[PubMed](#)]
- Mann, J.A.; MacDonald, B.A.; Kuo, I.; Li, X.; Broadbent, E. People respond better to robots than computer tablets delivering healthcare instructions. *Comput. Hum. Behav.* **2015**, *43*, 112–117. [[CrossRef](#)]
- VanLehn, K. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educ. Psychol.* **2011**, *46*, 197–221. [[CrossRef](#)]
- Brown, L.; Kerwin, R.; Howard, A.M. Applying behavioral strategies for student engagement using a robotic educational agent. In Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 13–16 October 2013; pp. 4360–4365. [[CrossRef](#)]
- Hattie, J. The applicability of Visible Learning to higher education. *Scholarsh. Teach. Learn. Psychol.* **2015**, *1*, 79–91. [[CrossRef](#)]
- Tiberius, R.G.; Billson, J.M. The social context of teaching and learning. *New Dir. Teach. Learn.* **1991**, 67–86. [[CrossRef](#)]
- Saerbeck, M.; Schut, T.; Bartneck, C.; Janse, M.D. Expressive robots in education: Varying the degree of social supportive behavior of a robotic tutor. In Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI '10), Atlanta, GA, USA, 10–15 April 2010; ACM: Atlanta, GA, USA, 2010; pp. 1613–1622. [[CrossRef](#)]
- Frymier, A.B.; Houser, M.L. The teacher-student relationship as an interpersonal relationship. *Commun. Educ.* **2000**, *49*, 207–219. [[CrossRef](#)]
- Skinner, E.A.; Belmont, M.J. Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *J. Educ. Psychol.* **1993**, *85*, 571–581. [[CrossRef](#)]
- Hamre, B.K.; Pianta, R.C. Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Dev.* **2001**, *72*, 625–638. [[CrossRef](#)]
- Kory-Westlund, J.M.; Breazeal, C.L. A long-term study of young children's rapport, social emulation, and language learning with a peer-like robot playmate in preschool. *Front. Robot. AI* **2019**, *6*, 81. [[CrossRef](#)]
- Alves-Oliveira, P.; Tullio, E.D.; Ribeiro, T.; Paiva, A. Meet me halfway: Eye behaviour as an expression of robot's language. In *AAAI Fall Symposium Series*; AAAI: Menlo Park, CA, USA, 2014; pp. 13–15.
- Atkinson, R.K.; Mayer, R.E.; Merrill, M.M. Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemp. Educ. Psychol.* **2005**, *30*, 117–139. [[CrossRef](#)]
- Boucher, J.D.; Pattacini, U.; Lelong, A.; Bailly, G.; Elisei, F.; Fagel, S.; Dominey, P.F.; Ventre-Dominey, J. I reach faster when I see you look: Gaze effects in human-human and human-robot face-to-face cooperation. *Front. Neurobot.* **2012**, *6*, 1–11. [[CrossRef](#)] [[PubMed](#)]
- Esposito, J. Negotiating the gaze and learning the hidden curriculum: A critical race analysis of the embodiment of female students of color at a predominantly white institution. *J. Crit. Educ. Policy Stud.* **2011**, *9*, 143–164.
- Konijn, E.A.; Smakman, M.; Van den Berghe, R. Use of robots in education. In *The International Encyclopedia of Media Psychology*; van den Bulck, J., Sharrer, E., Ewoldsen, D., Mares, M.-L., Eds.; Wiley: New York, NY, USA, 2020.
- Konijn, E.A.; Hoom, J.F. Robot tutor and pupils' educational ability: Teaching the times tables. *Comput. Educ.* **2020**, *157*, 103970. [[CrossRef](#)]
- Kennedy, J.; Baxter, P.; Senft, E.; Belpaeme, T. Social robot tutoring for child second language learning. In Proceedings of the 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Christchurch, New Zealand, 7–10 March 2016; pp. 231–238. [[CrossRef](#)]
- Vogt, P.; Haas, M.D.; Jong, C.D.; Baxter, P.; Krahmer, E. Child-Robot Interactions for Second Language Tutoring to Preschool Children. *Front. Hum. Neurosci.* **2017**, *11*, 73. [[CrossRef](#)]

25. Kory-Westlund, J.M.; Jeong, S.; Park Hae, W.; Ronfard, S.; Adhikari, A.; Harris, P.L.; DeSteno, D.; Breazeal, C.L. Flat versus expressive storytelling: Learning and retention of a robot's narrative. *Front. Hum. Neurosci.* **2017**, *11*, 643–648. [[CrossRef](#)]
26. Hindriks, K.V.; Liebens, S. A robot math tutor that gives feedback. In *Social Robotics. ICSR 2019. Lecture Notes in Computer Science*; Salichs, M., Ge, S.S., Barakova, E.I., Cabibihan, J.-J., Wagner, A.R., Castro-González, Á., He, H.S., Eds.; Springer: Cham, Switzerland, 2019; Volume 11876, pp. 601–610. [[CrossRef](#)]
27. Kennedy, J.; Baxter, P.; Belpaeme, T. The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, Portland, OR, USA, 2–5 March 2015; ACM: New York, NY, USA, 2015; pp. 67–74. [[CrossRef](#)]
28. Wei, C.W.; Hung, I.C.; Lee, L.; Chen, N.S. A joyful classroom learning system with robot learning companion for children to learn mathematics multiplication. *Turk. Online J. Educ. Technol.* **2011**, *10*, 11–23.
29. Huang, I.S.; Hoorn, J.F. Having an Einstein in class. Teaching maths with robots is different for boys and girls. In *Proceedings of the 13th World Congress on Intelligent Control and Automation (WCICA 2018)*, Changsha, China, 4–8 July 2018; Wang, X., Wang, Z., Wu, J., Wang, L., Eds.; IEEE: New York, NY, USA, 2018; pp. 424–427. Available online: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8630584> (accessed on 25 October 2020). [[CrossRef](#)]
30. Krämer, N.C.; Karacora, B.; Lucas, G.; Dehghani, M.; Rütther, G.; Gratch, J. Closing the gender gap in STEM with friendly male instructors? On the effects of rapport behavior and gender of a virtual agent in an instructional interaction. *Comput. Educ.* **2016**, *99*, 1–13. [[CrossRef](#)]
31. Arroyo, I.; Royer, J.M.; Park Woolf, B. Using an intelligent tutor and math fluency training to improve math performance. *Int. J. Artif. Intell. Educ.* **2011**, *21*, 135–152. [[CrossRef](#)]
32. Konijn, E.A.; Hoorn, J.F. Media psychological perspectives on the use of communication robots in health care. In *The International Encyclopedia of Media Psychology*; van den Bulck, J., Ed.; Wiley: New York, NY, USA, 2020. [[CrossRef](#)]
33. Jamone, L.; Ugur, E.; Cangelosi, A.; Fadiga, L.; Bernardino, A.; Piater, J.; Santos-Victor, J. Affordances in psychology, neuroscience, and robotics: A survey. *IEEE Trans. Cogn. Dev. Syst.* **2018**, *10*, 4–25. [[CrossRef](#)]
34. Syrdal, D.S.; Dautenhahn, K.; Woods, S.N.; Walters, M.L.; Koay, K.L. Looking good? Appearance preferences and robot personality inferences at zero acquaintance. In *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*; AAAI: Menlo Park, CA, USA, 2007; pp. 26–28.
35. Köse, H.; Uluer, P.; Akalın, N.; Yorgancı, R.; Özkul, A.; Ince, G. The effect of embodiment in sign language tutoring with assistive humanoid robots. *Int. J. Soc. Robot.* **2015**, *7*, 537–548. [[CrossRef](#)]
36. Moshkina, L.; Trickett, S.; Trafton, J.G. Social engagement in public places: A tale of one robot. In *Proceedings of the 2014 ACM/IEEE international conference on human-robot interaction (HRI '14)*, Bielefeld, Germany, 3–6 March 2014; ACM: New York, NY, USA, 2014; pp. 382–389. [[CrossRef](#)]
37. Li, D.; Rau, P.L.P.; Li, Y. A cross-cultural study: Effect of robot appearance and task. *Int. J. Soc. Robot.* **2010**, *2*, 175–186. [[CrossRef](#)]
38. Paauwe, R.A.; Hoorn, J.F.; Konijn, E.A.; Keyson, D.V. Designing robot embodiments for social interaction: Affordances topple realism and aesthetics. *Int. J. Soc. Robot.* **2015**, *7*, 697–708. [[CrossRef](#)]
39. Van Vugt, H.C.; Konijn, E.A.; Hoorn, J.F.; Eliëns, A.; Keur, I. Realism is not all! User engagement with task-related interface characters. *Interact. Comput.* **2007**, *19*, 267–280. [[CrossRef](#)]
40. Konijn, E.A.; Hoorn, J.F. Parasocial interaction and beyond: Media personae and affective bonding. In *The International Encyclopedia of Media Effects*; Roessler, P., Hoffner, C., van Zoonen, L., Eds.; Wiley-Blackwell: New York, NY, USA, 2017; pp. 1–15. [[CrossRef](#)]
41. Chen, H.; Park Hae, W.; Breazeal, C.L. Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Comput. Educ.* **2020**, *150*, 103836. [[CrossRef](#)]
42. Van Vugt, H.C.; Hoorn, J.F.; Konijn, E.A.; De Bie Dimitriadou, A. Affective affordances: Improving interface character engagement through interaction. *Int. J. Hum. Comput. Stud.* **2006**, *64*, 874–888. [[CrossRef](#)]
43. Onyeulo, E.B.; Gandhi, V. What makes a social robot good at interacting with humans? *Information* **2020**, *11*, 43. [[CrossRef](#)]
44. Beckmann, E.; Minnaert, A. Non-cognitive characteristics of gifted students with learning disabilities: An in-depth systematic review. *Front. Psychol.* **2018**, *9*, 504. [[CrossRef](#)]
45. Serholt, S.; Barendregt, W. Robots tutoring children: Longitudinal evaluation of social engagement in child-robot interaction. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*, Gothenburg, Sweden, 23–27 October 2016; pp. 1–10. [[CrossRef](#)]
46. Imai, M.; Ono, T.; Ishiguro, H. Physical relation and expression: Joint attention for human-robot interaction. In *Proceedings of the 10th IEEE International Workshop on Robot and Human Interactive Communication. ROMAN 2001 (Cat. No.01TH8591)*, Bordeaux, Paris, France, 18–21 September 2001; Volume 50, pp. 636–643. [[CrossRef](#)]
47. Mutlu, B.; Forlizzi, J.; Hodgins, J. A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Proceedings of the 2006 6th IEEE-RAS International Conference on Humanoid Robots*, Genova, Italy, 4–6 December 2006; pp. 518–523. [[CrossRef](#)]
48. Mader, J. How Teacher Training Hinders Special-Needs Students. *The Atlantic Daily*. Available online: [www.theatlantic.com/education/archive/2017/03/how-teacher-training-hinders-special-needs-students/518286/](http://www.theatlantic.com/education/archive/2017/03/how-teacher-training-hinders-special-needs-students/518286/) (accessed on 3 March 2017).
49. Lee, F.; Yeung, A.; Tracey, D.; Barker, K. Inclusion of children with special needs in early childhood education. *Top. Early Child. Spec. Educ.* **2015**, *35*, 79–88. [[CrossRef](#)]

- 
50. Leyzberg, D.; Spaulding, S.; Toneva, M.; Scassellati, B. The physical presence of a robot tutor increases cognitive learning gains. In Proceedings of the Annual Meeting of the Cognitive Science Society, Sapporo, Japan, 1–4 August 2012; Volume 34, pp. 1882–1887. Available online: <https://escholarship.org/uc/item/7ck0p200> (accessed on 17 June 2020).
  51. Smakman, M.; Konijn, E.A. Robot tutors: Welcome or ethically questionable? In *Robotics in Education, Advances in Intelligent Systems and Computing (AISC), 1023*; Merdan, M., Lepuschitz, W., Koppensteiner, G., Balogh, R., Obdržálek, D., Eds.; Springer: Cham, Switzerland, 2020; pp. 376–386. [[CrossRef](#)]